文献与数据学报
第 6 卷　第 4 期　　2024 年 12 月

青年用户申贷行为的群体画像构建研究

——来自手机银行 APP 的大样本数据*

窦路遥
（中国科学院武汉文献情报中心，武汉　430071）

摘　要:［目的 / 意义］挖掘青年用户申贷理由文本的主题特征，构建用户画像标签体系，有助于系统理解青年用户线上手机银行申贷行为的偏好规律，从而为手机银行平台适龄化建设提供优化策略。［方法 / 过程］收集手机银行 APP 后台大规模、高精度的青年用户申贷数据共 468 647 条，利用 BERTopic 模型凝练申贷理由主题，基于申贷行为特点从财务属性、人口属性、职业属性、生活属性、信用属性 5 个维度抽取画像标签，结合银行用户评级赋分标准构建用户画像识别标签体系，通过 K-means 聚类实现用户画像，呈现群组特征并进行差异对比分析。［结果 / 结论］研究识别出青年用户的申贷理由共计 30 个主题，涵盖日常生活、财务管理、资产支出、个人兴趣、家庭生活和生产经营六个行为层面类别；根据这些主题和属性特征，将青年用户申贷群体细分为生产经营型、逾期借贷型、申请备用型、消费导向型和初次申请型五类。因此，各青年用户群体画像之间存在显著差异，建议采取针对性的激励措施，增强手机银行用户的黏性和满意度。

关键词: 手机银行　青年用户　申贷行为　用户画像
分类号: G358　F832
DOI: 10.31193/SSAP.J.ISSN.2096-6695.2024.04.08

0 引 言

随着数字化技术的进步和人工智能技术的发展，手机银行 APP 逐渐成为现代社会生活中较为关键的金融服务平台。手机银行 APP 作为银行数字化转型的重要抓手，在金融促活、普惠信贷、科技金融层面上发挥着不可或缺的作用。党的二十大报告明确要求加强金融工具创新，强化金融机构自身数据能力建设，加强开发增进民生福祉的线上金融产品。在这一背景下，用户申贷

* 本文系国家自然科学基金项目“多源数据融合场景下的对抗式隐私洞察靶向保护技术研究”（项目编号：61902226）研究成果之一。
［作者简介］窦路遥（ORCID：0009-0005-5075-0613），男，博士生，研究方向为用户信息行为，Email：douly@mail.whlib.ac.cn。

信息平台作为普惠金融工具创新的重要窗口，就成为促进数据要素流通和数据确权定价的落脚点，这也是我国加快建设“数据强国”与“金融强国”的关键路径。

当前，个人用户申贷逐渐由传统线下审批转向线上大数据平台评估，个人信息集合、申请理由文本、具体资产状况等核心要素已成为判断申请结果的重要情境。其中，不同年龄群体所具备的信息特征属性和申贷理由存在较大差异，而熟练掌握线上金融工具操作的青年用户在申贷阶段所提供的行为信息兼具完整性和导向性。依托线上大数据平台更方便快捷且更具隐私安全的申请环境，青年用户申贷行为表现亦呈现出信息多元化、目的特征化、群体个性化的融合趋势[1]。由此可见，青年用户申贷过程中所蕴含的结构化行为信息较之其他年龄段的用户更具识别性。然而，融入青年用户金融习惯的复杂信贷信息仍面临着申请信息杂糅和申请理由模糊的问题，同时，不同青年群体的金融行为和需求可能存在显著差异，导致金融机构无法全面揭示用户申贷行为与申请偏好。为此，现阶段亟须拓展申贷理由内容差异与群体识别等领域的研究。

鉴于此，本文基于手机银行APP的真实申贷数据，对青年用户申贷理由进行主题挖掘，结合用户评级赋分标准和申贷标签明细构建青年用户画像标签体系，进行群体聚类并抽象出不同用户群体的典型特征，最后通过对比分析揭示不同群体间的画像差异，实现青年用户申贷行为的系统刻画。旨在通过研究建立青年用户画像标签体系，实现群体聚类与特征抽象并进行青年用户申贷行为的多维度分析，这有助于差异化金融产品精准推送和分类营销，从而推动用户申贷行为信息建设的适龄化发展和创新化改进，进一步为金融产品的创新发展提供参考。

1 相关研究

1.1 青年用户申贷行为信息

青年用户申贷行为指的是青年人群在金融机构申请信贷产品的行为，行为信息包括借款需求、借款目的、贷款偏好等信息特征[2-3]。青年用户在申贷行为上具有高频小额贷款[4]、依赖数字平台[5]以及还款能力波动较大[6]的群体特征。其中，青年用户的申贷行为多集中于小额消费贷款，主要用于满足日常消费、教育、电子产品等需求；相比传统银行贷款，青年群体更倾向于通过数字平台（如互联网金融公司、P2P贷款平台）进行申贷；青年用户还款能力受职业发展和收入不稳定的影响较大，尤其是在刚步入职场的阶段，还款压力较为明显。尽管申贷行为给青年用户带来了信贷便利，但同时也伴随着诸多风险。例如，由于消费需求旺盛和还款能力不稳定，青年用户往往面临债务累积风险，容易陷入“以贷养贷”的恶性循环[7]。同时部分青年用户缺乏金融知识，容易被高利贷、非法借贷平台所吸引，导致债务危机[8]。而不良的申贷记录也会对青年用户的个人信用产生长期影响，影响未来的贷款资格和职业发展[9]。

1.2 用户画像

用户画像是建立在真实用户数据基础上的目标模型[10]，在信贷判别、健康社区、政府服务等领域主要用于用户需求识别、精准信息营销、平台个性化服务等工作，其构建方式一般基于用

户行为[11]、用户偏好[12]、文本主题[13]、用户情绪[14]等。现有关于用户申贷行为画像的研究主要依赖于用户的人口统计学指标（出生地、职业、年龄等）[15]、行为计数指标（产品使用次数、业务往来频率、资金流动性等）[16]以及社交网络指标（出入度、中心度、凝聚子群等）[17]，实现对用户群体的抽象和特征描述。近年来，也有部分学者尝试通过动态弹幕评论[18]、日志解析[19]、融合多源数据[20]的形式实现用户画像的精准构建。例如，杨阳等[18]以 B 站教学类视频为例，构建动态用户画像模型，探究用户发送弹幕与观看视频分集的行为随不同时间维度的变化情况；李嘉兴等[19]基于移动终端日志构建微信老年用户使用行为画像；严炜炜等[20]通过对新浪微博、百度贴吧、CSDN 社区的信息整合，构建用户知识交流主题，分析多平台用户画像和知识交流主题的差异。

1.3 研究述评

伴随申贷数据的几何式增长，大数据平台所收集的用户申贷信息表现出明显的用户特征杂糅性和申请理由主观性等问题。同时，用户画像技术虽可以应用于不同场景以揭示用户特征规律，但现有研究大多依赖于理论层面上的相关指标和针对性涣散的网络数据，而申请信息准确聚焦、申请理由充实全面的样本研究较少。尤其是基于大规模、高精度、贴实际的青年用户申请数据所进行的画像较为有限，缺乏对青年用户申贷行为情境下的用户整体画像揭示，因而未能深入理解青年用户申贷需求的全貌特征与偏好。

2 研究设计

青年用户作为数据要素市场的重要组成部分，其申贷行为信息蕴含着数据要素流通趋势的变化方向，以及青年新颖的消费习惯和金融需求。与传统的青年概念相比，信贷领域中的青年概念更强调信贷认知与还款能力。在信贷实践中，18 岁左右的个体由于尚未形成足够的社会和金融资本，通常难以满足银行的信贷审批标准。本科生一般在 22~23 岁毕业，而信贷要求往往包括社保或公积金的缴纳记录，这通常需要至少 1 到 2 年的时间积累，使得 25 岁成为较为合理的信贷行为发生起点。同理，接近 50 岁的群体在信贷申请中受限于退休年龄和子女已具备借贷能力等因素，银行对其信贷需求的审查也更为严格，进一步降低了该群体的信贷行为活跃度，使得 45 岁人成为较为合理的信贷行为发生终点。因此，本文的研究对象为青年申贷用户，对青年用户的年龄范围限制在 25 岁至 45 岁。

手机银行 APP 的出现为青年用户申贷提供了优化用户体验、标准化数据收集流程、提高实时性及系统化管理方面的便利，其格式化的信息上传流程也为系统化的数据收集奠定了基础[21]。本文基于某手机银行 APP 后台的青年用户申贷信息作为样本数据进行研究，研究框架如图 1 所示。首先，提取 APP 后台的非结构化申请数据并进行预处理；其次，在实验设计部分，利用 BERTopic 模型对青年用户申贷理由进行主题提取，构建申贷核心主题，筛选出申请理由明确、内容详细且可操作的青年用户数据，并对该部分用户构建用户画像标签体系；最后，在青年用户画像呈现部分，采用聚类算法生成用户画像，并分析不同用户画像间的构型差异。

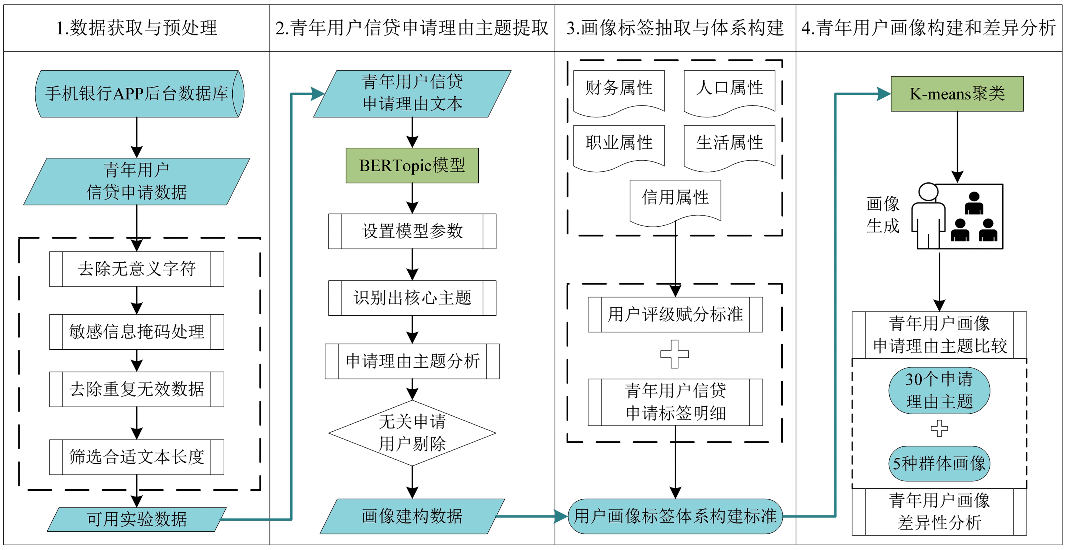


图 1　研究框架图

2.1　数据来源与预处理

手机银行是指通过手机终端接入银行金融服务系统，并进行资金转移、信息查询、金融交易等操作的一种金融服务形式[22]。通过手机银行应用，用户可以查看其账户余额、交易历史、信用卡账单等信息，帮助用户评估自己的财务状况。基于这些信息，用户可以决定是否需要借贷资金来满足特定的需求，或者通过填写表格、上传必要文件等方式提交在线信用贷款申请。

本文基于某手机银行 APP 后台信息，将青年用户在 APP 中的申贷行为记录及其关联的历史数据进行匿名化和脱敏处理，获取客户在 2021~2023 年间的申贷数据共 681 166 条（已留存备索）。该申请数据囊括所在省 16 个地级市的手机银行借贷申请信息，申请人类型包括农户与非农户、公积金客户与社保客户、有营业执照型客户和无营业执照型客户；贷款申请方式主要为信用贷款。考虑到数据可用性，对 681 166 条原始数据进行预处理，包括：去除申请信息中的无意义字符，对个人敏感信息进行掩码处理；选择年龄范围在 25 岁至 45 岁间的青年用户作为研究对象；去除重复申请数据和申请无效数据；筛选出申请理由文本长度在 5~500 之间的数据作为可用数据。经过数据预处理之后，共得到可用申请数据 468 647 条。数据示例见图 2。

2.2　青年用户申贷理由文本的主题提取

用户申贷理由作为评价用户申请意愿、分析申请可信度、拟定借款还款方式的关键文本内容之一[23]，其中富含的用户行为和申请逻辑具有非常强的上下文关联和长距离依赖关系（申请理由文本简要示例见图 3）。为此，本文采用 BERTopic 模型来挖掘青年用户申贷理由主题。

BERTopic 是一种基于变换器的无监督主题聚类深度学习模型，在处理大规模文本数据时具有出色性能[24]，因此与本研究的实验数据规模相匹配。该模型的详细实现步骤如下：首先，针对上文收集到的 468 647 条可用申请数据，利用 jieba 分词工具将每条申请数据实现分词处理，得到一个词汇列表；其次，采用 BERT 模型进行词嵌入处理；再次，由于高维度向量空间的复杂

性，采用UMAP算法对文档向量进行降维，随后使用HDBSCAN算法识别具有不同密度的聚类簇，并自动检测与过滤噪音数据点[25]；最后，采用 c-TF-IDF 模型实现主题提取[26]。

客户编号	是否涉农	担保方式	贷款金额（元）	贷款利率（%）	核心申请理由
***6874348**	是	信用	80000	6.09	购买车辆：作为一名销售人员，本人需要频繁外出拜访客户或参加行业会议。由于公共交通工具的不稳定性，最近一次本人因为公交车晚点而错过了一个重要的商务会议。为了提高工作效率，减少因公共交通工具造成的时间浪费，拥有一辆经济型轿车将极大地提高工作灵活性和效率。预计购买车辆的预算为10万元，希望贷款金额为8万元，剩余部分通过存款支付。本人计划选择较低利率的贷款产品，分期三年偿还，每月还款金额不超过3000元。除了日常拜访客户和参加会议外，车辆还将用于参加公司组织的外地业务拓展活动。
***4137570**	是	信用	140000	7.83	健康治疗费用：最近，本人的父母都出现了健康问题，需要进行长期的治疗和护理。父亲被诊断出患有慢性心脏病，母亲则需要进行膝关节置换手术。这些医疗费用和药品费用相当昂贵，而本人目前的收入无法完全覆盖这些支出。预计父亲的心脏病治疗费用每年需要5万元，母亲的手术和康复费用总计约为15万元。因此，本人计划贷款14万元，用于支付父母未来两年的医疗费用，确保他们能够得到及时的治疗和照顾。希望选择长期还款计划，分期五年偿还，每月还款金额不超过4000元。此外，本人计划为父母购买医疗保险，以减轻未来可能产生的额外医疗费用。
***4215656**	否	信用	300000	7.39	扩大生意规模：本人经营着一家小型制造企业，主要生产环保包装材料。公司的产品质量和客户口碑一直很好，市场需求也在不断增长。为了满足市场需求并提高企业竞争力，本人计划扩大生意规模，增加生产线和扩建厂房。这次扩展计划包括新建一条全自动生产线，增加厂房面积500平方米，预计总投资为150万元。希望贷款30万元，其余部分通过企业自有资金和股东注资解决。计划选择中长期还款计划，分期五年偿还，每月还款金额不超过0.8万元。预计扩展后年营业额将增加30%，净利润提高20%，三年内可收回投资成本。贷款将是实现这个发展目标的关键支持，帮助本人进一步拓展企业业务，提升经济效益。

图 2　信贷申请数据示例

XXX，身份证号码XXXXXXXXXXXX，29岁，XX区XX镇XXX人，目前就职于碧桂园服务股份有限公司。其丈夫XXX，身份证号码XXXXXXXXXXXX，30岁，XX县XX镇XXX人，主要从事服装服饰、鞋帽、箱包、饰品、日用百货零售，其经营实体为XX市XXX服饰有限公司，该经营实体目前为双方共同经营。经营公司2020创立，位置处于XX市XX区吾悦广场2层 XX号铺，商铺位于闹市人流聚集处，经营状况良好。现因想增加货物储备扩大经营规模、提升商品品质，导致流动资金不足，为满足当前经营需要，故申请信用贷款。贷款期间可能存在的风险点在于：收入还款来源较为单一，若遇经营变动，收入下降，可能会影响贷款收回，从事服装行业，受季节波动影响较大。借款人家庭资产情况：轿车一辆，价值约10万元；资产合计 10万元借款人负债情况：借款人家庭负债主要为我行经营性贷款16万元；负债合计16万元。借款人信用记录良好，其配偶存在2次征信纪录逾期；夫妻二人无不良嗜好。申请人XXX信用等级评定为B级，拟申请借款16万元，期限1年，拟执行“惠民贷款”利率5.5%，用于经营服装行业使用，支付方式为自主支付，担保方式为信用，还款方式为按季结息，到期还本。

图 3　申请理由文本简要示例

2.3　画像标签抽取与体系构建

青年用户在手机银行 APP 所进行的申贷操作需要填写相应的关键信息，如基础类信息（年龄、单位、性别等）、历史信用状况信息（借款记录、征信情况、有无担保等）、资金类信息（个人资产、流动资产、不动产等）等。本文结合这些关键信息及已有研究结果[27-28]，对青年用户申贷信息的特征进行高度凝练，并设置出相应的画像标签用于构建用户画像模型。

现有研究多从财务属性、人口属性、职业属性、生活属性、信用属性等维度抽取画像标签[29-31]。财务属性主要考察用户的资产水平、偿债水平及既存贷款水平；人口属性则聚焦于用户的性别、年龄、婚配、身体、学历层面；职业属性不仅考虑用户所在行业和具体实操职业，也量化评估生产经营的时间水平；生活属性一般考虑用户的属地情况和家庭水平；信用属性重点关

注用户自身的信用风险及其延伸因素。结合多因素评估视角以及青年用户申贷行为特点，将类似要素进行归纳整合，构建青年用户申贷画像标签体系，如表 1 所示。

表 1　青年用户申贷画像标签体系

属性	标签	属性说明
财务属性	总资产、资产负债率、偿债能力、贷款占用形态	评估个人或家庭的财务健康状况，包括资产规模、资产结构、负债水平等，反映申请用户的财务风险和偿债能力
人口属性	性别、年龄、婚姻状况、健康状况、最高学历	考虑到不同人口特征对用户经济行为和风险承担的影响，如年龄、性别、健康状况等会影响收入水平、负债能力等
职业属性	行业、职业、工作时间、生产经营能力、自办经济实体、生产经营年限	考察用户的职业稳定性、收入水平和社会地位，不同职业领域的收入水平和稳定性存在显著差异
生活属性	在本地居住时间、是否本地人、居住状况、家庭人均月收入、户主生活方式	考虑到用户的生活稳定性和保障程度，包括居住情况、社会保障、人均收入等
信用属性	本行信用记录、他行信用记录、本行业务使用情况、贷款按期付息情况、对外担保	考察用户的信用记录和信用行为，反映用户的信用状况和信用风险水平

在明确画像标签体系之后，结合用户评级赋分标准和青年用户申贷标签明细，整合相似部分，制定出用户画像标签体系构建标准（见图 4）。基于图 4 对用户画像标签的“L8 行业”与“L9 职业”进行具体分类，如图 5 所示。

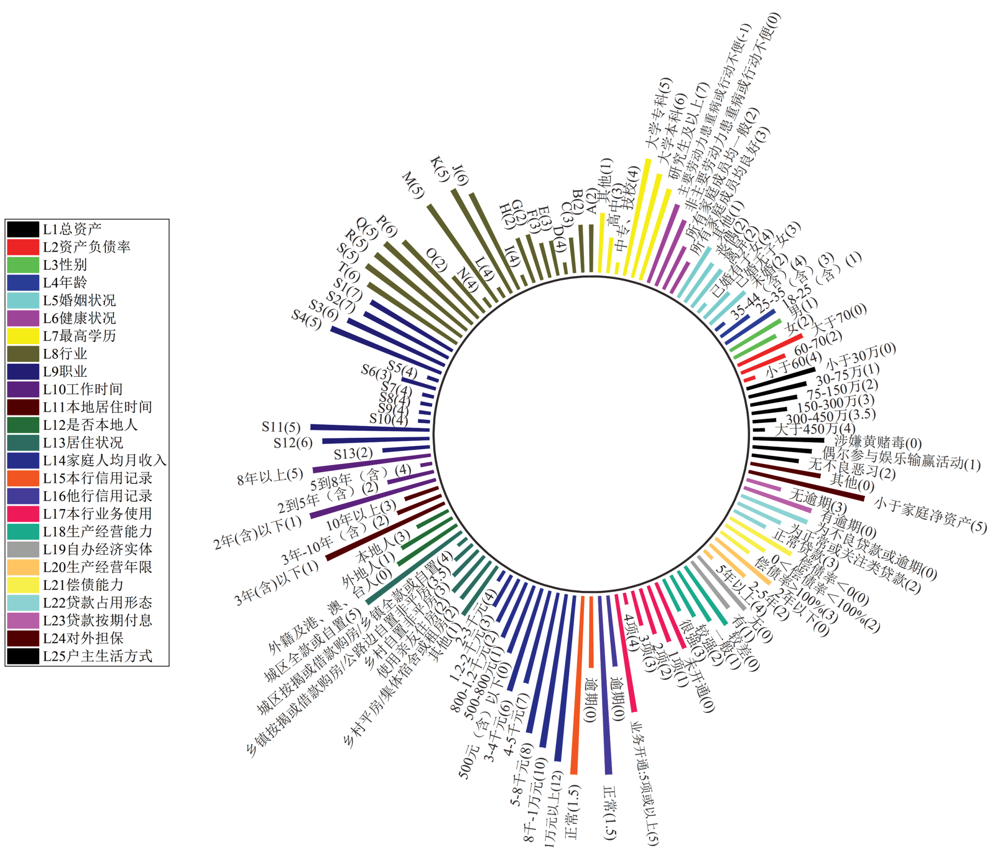


图 4　用户画像标签体系构建标准

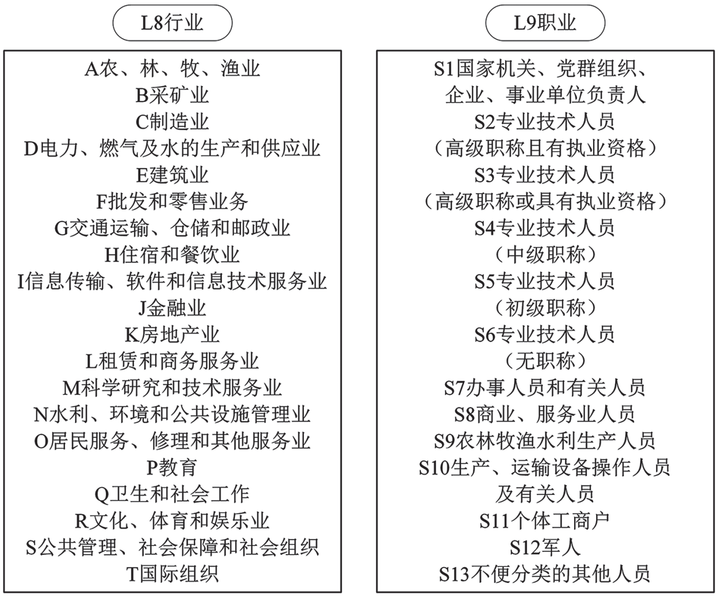


图 5　用户画像标签中“L8 行业”和“L9 职业”的具体分类

3　实证分析

3.1　青年用户申贷理由的主题提取

对 BERTopic 模型进行调参训练，迭代生成 30 个具备明确含义的主题（详见表 2）。该结果解释了 461 166 个申贷理由的主题分类表征，另有 7 481 个申贷理由因为主题杂糅或离群被视为噪声值而剔除，离群表征主要包括申请理由违背公序良俗（如赌博、走私）、国家政策限制（如信贷资金流入房市、股市、债市）以及资金意图违规（如“垒大户”、高利贷）。

表 2　用户申贷理由的主题提取

占比（%）	类别	主题命名	部分关键词
19.60	日常生活	家庭日常开销	日用品；水电；杂货开销；生活费；食材；家具；生活用品；家庭开支
		生活用品采购	采购；家居；必需品；生活必备；生活用具；品质；家庭物品；生活产品
		租房住宿费用	租房；费用；外出住宿；房租；租金；二手住房；出租；间断居住
		生活用品采购	采购；家居；必需品；生活必备；生活用具；品质；家庭物品；生活产品
		租房住宿费用	租房；费用；外出住宿；房租；租金；二手住房；出租；间断居住
		个人娱乐消费	娱乐；消费；休闲；活动；娱乐项目；游戏；电影；放松；玩乐；个人
		社交活动费用	社交活动费用；朋友聚会；人情花销；部门聚餐；宴会宴席；招待费
		休闲活动	渔具购买；刺激活动；文娱节目；游乐；器材消费；娱乐项目；场所
		急救医疗支出	急救品；医疗费；急救车；医疗支出；慢性病药品；诊疗费；护理费
16.07	财务管理	投资规划	基金投资；理财购买；收益型产品；高中低风险；资产规划
		保险费	保险合同；两全；年金；保费；保障；理赔；养老；投保；保险费
		学杂费	书本费；补习费；教育额外支出；培训费；孩子；教材；学生
		健康保健	保健活动；健身训练；体检；保健品；药品；保健费；保健产品

续表

占比（%）	类别	主题命名	部分关键词
13.71	资产支出	交通工具购买	车辆购买；买车；车款；品牌；型号；性能；价格；二手车；出行；保养
		汽车保养	保养费支出；检修；汽车维修；零配件；机油；轮胎；车灯；车身；车窗
		房屋装修	木质材料；设计费用；装饰；预算；施工人工；装修费用；装潢服务
		电器家具	洗衣机；抽烟机；家具；电器产品；冰箱；空调；席梦思；四件套
9.31	个人兴趣	职业技能提升	课外培训；考试；学习；技能；职场；提升；课程；进修；认证；专业
		数码产品购买	数码设备；智能手机；照相机配件；索尼；电脑市场；无人机；平板
		家居装饰购买	地板装饰；打柜；家具；布置；家居配饰；风格设计；平面图
		宠物医疗费用	宠物洗澡；宠物医疗费用；打虫药；抗生素；小型手术；疫苗；检查
		美容美发	美容；美发；护理；美甲；美容院；理发；化妆品；美发产品；美容产品
9.85	家庭生活	庆典活动	生日升学庆典；摆设花费；策划；装饰；餐饮；场地；物料；服务；支出
		假日出行	假日；远途油费；度假；旅游；活动；家庭；休闲；出游；花费；旅行支出
		婚庆费用	婚礼搭建；策划；布置；婚宴；装饰；婚庆费；礼金；司仪；结婚
		家居设备更新	更新；设备；智能家居；家电检查；家具；燃气更换；器具；零件采购
31.46	生产经营	食品果蔬生产	花生糖；罐头；苹果；火龙果；蓝莓；脆桃；奶油草莓；8424 西瓜
		服装加工	男装；女装；中性装；棉质；麻质；毛质；丝质；晚装；冬装；春秋装
		产品制造	数控机床；加工中心；注塑机；硫化机；塔吊；升降机；搅拌机
		商品贸易	零食；香烟；白酒；红酒；茶叶；瓜片；石斛；米斛；玩具；鲜花
		水产养殖	龙虾；鱼塘；鱼苗；黄鳝；挖掘；水泵；稻鱼一体；水鸭；供氧器
		农作物耕种	水稻；旋耕机；稻种；化肥；杀虫剂；除草剂；氮肥；复合肥；秋收

依据用户申贷行为和主题分类的相关研究，将 30 个主题划分为日常生活、财务管理、资产支出、个人兴趣、家庭生活、生产经营 6 个类别（详见表 2，占比表示该类别所包含主题占据全部主题的比例）。结果显示，生产经营类别最受青年用户关注，包括食品果蔬生产、服装加工、产品制造、商品贸易、水产养殖、农作物耕种，其中的农作物耕种占比最高，达到了 9.59%，说明青年用户的借贷心理预期主要聚焦在资金的可持续利用上。此外，青年用户申贷理由也倾向于日常生活中的家庭日常开销和生活用品采购，以及资产支出类别上的交通工具购买，反映出青年用户对于信用贷款的态度部分聚焦于生活消费以及交通工具使用方面。

3.2　标签体系分类统计

根据用户画像标签体系构建评分规则，对 461 166 个手机银行个人用户的标签进行赋值。为了排除标签间相关性对最终聚类结果的影响，采用相关分析法对用户画像标签进行相关性检验（结果详见图 6）。结果显示大多数标签之间未出现显著相关性，多数为不相关或弱相关（0~0.4），表明标签呈现较强的独立性[32]。因此，上文所构建的标签体系适用于后续的聚类分析。

对构建的用户画像标签体系进行分类统计，其标签分类情况见图 7，详细数值已留存备索。其中，超过一半的青年用户申请者资产负债率小于 60，而总资产的分布呈现出明显的倾斜，约有 32% 的申请者资产分布在 150~300 万元之间。男性申请者占据绝大多数，占比达到 70.12%；年龄分布则呈现出以 25~35 岁为主的趋势，占比达到 33.33%。以办事人员和商业服务业人员为主

的申请群体，占比分别为 14.80% 和 18.71%，而申请者的家庭人均月收入主要集中在 4000~5000 元和 5000~8000 元两个区间，占比分别为 44.54% 和 23.74%。

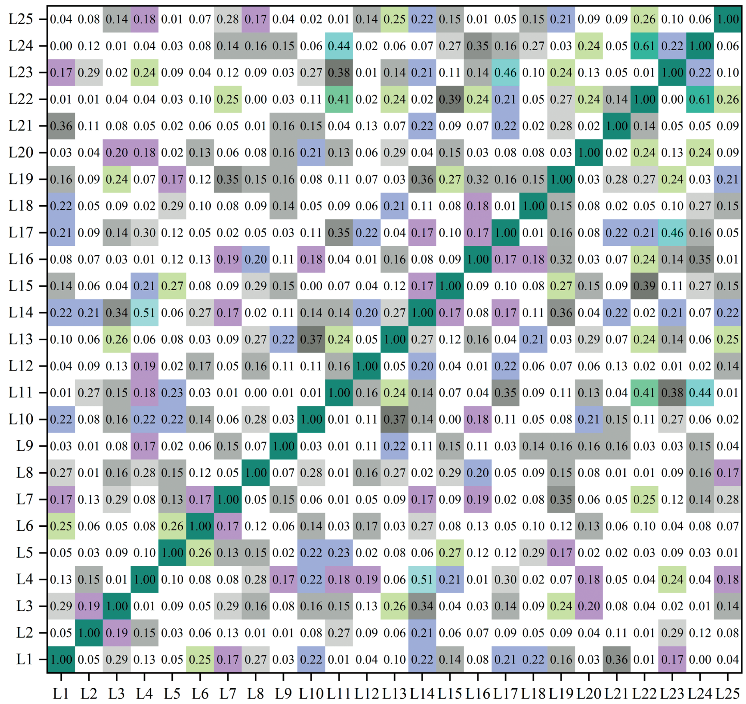


图 6　标签相关性检验

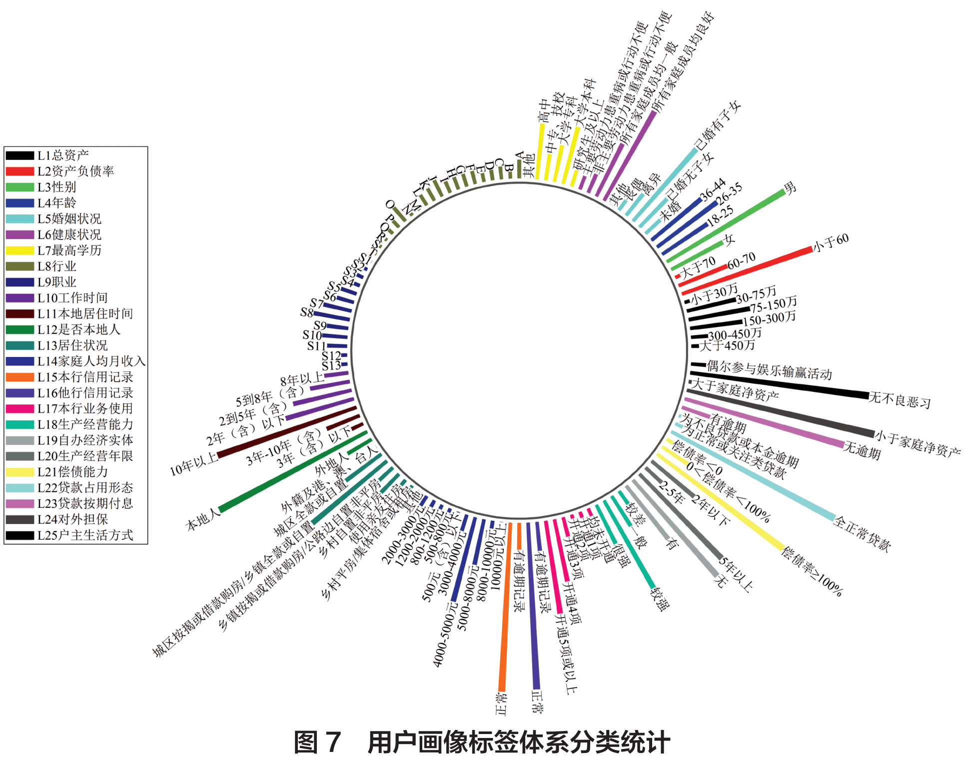


图 7　用户画像标签体系分类统计

注：图中柱线长度表示各用户画像标签分类的数值占比。

4 用户画像描述与结果讨论

剔除噪声数据后，对识别出明确主题的461 166名青年用户申贷者进行画像分析。先将25类标签化的纯数值数据进行主成分降维成6个核心变量，使用上文所提到的UMAP算法将6维数据降维到2维，再采用K-means算法对用户画像标签体系进行群体聚类。通过手肘法选取最佳聚类数K，发现当K取值小于5时和方差下降明显，往后再增加K所得的聚合效果回报迅速变小，因此确定最佳聚类值为5[33]。依据群体标签特点，本文对聚类所得的5类群体分别命名为生产经营型用户、逾期借贷型用户、申请备用型用户、消费导向型用户、初次申请型用户。这5种类型用户的申请理由主题详见表3。

表3 各类型青年用户的申请理由主题

用户画像分类	申请理由主题
生产经营型用户	产品制造、水产养殖、农作物耕种
逾期借贷型用户	个人娱乐消费、投资规划、汽车保养
申请备用型用户	急救医疗支出、保险费、婚庆费用
消费导向型用户	租房住宿费用、个人娱乐消费、休闲活动
初次申请型用户	健康保健、电器家具、假日出行

上述不同类型用户群体的申请理由主题反映出他们在资金需求和消费行为上的差异。比如，生产经营型用户主要关注生产经营活动，而消费导向型用户则更注重个人消费和生活享受。同时，逾期借贷型用户的逾期行为产生原因可能倾向于个人高额消费以及投资亏损，而申请备用型用户的申请理由则趋向于大众化事件，如医疗、保险、婚姻等。此外，不同群体的职业和行业分布存在差异，比如，生产经营型用户可能更多从事农业或制造业等行业，而消费导向型用户可能更多从事服务业或娱乐业等行业。下面分别详细分析与描述。

4.1 生产经营型用户

生产经营型用户在整个用户画像群体中占比最高，达到35.82%。这一群体申请信贷的主要原因是为了商业化经营和产品生产，其财务属性表现稳健，资产负债率维持在可控范围之内，偿债能力和贷款情况普遍良好。从人口和职业属性来看，该群体以男性为主导，且已婚有子女的比例相对较高，年龄分布在整个青年阶段较为均匀，表明其在不同年龄段都有较强的商业活力和经营动力。其主营业务涵盖批发和零售、住宿和餐饮、农林牧渔等多个领域，展现出在本地经济活动中的重要性及在满足市场需求方面的多样性和灵活性。

此外，该群体对本地社会环境和市场背景具备一定认识，能够较好地适应和把握本地市场变化。拥有自办经济实体的比例高达100%，表明他们不仅是积极的市场参与者，更是当地经济活动的推动者和创新者。在持续经营过程中，该群体的信用记录良好，表现出较高的信用意识和责任感，并能够积极融入手机银行系列业务，利用现代金融科技手段提升经营效率和管理水平。综合来看，生产经营型用户主要为本地年轻生意人，他们经济收入尚可，银户关系和谐，这使得他们成为青年用户申贷行为的重要组成部分。

4.2 逾期借贷型用户

该群体占青年用户总量的六分之一，申请信贷的理由偏向续贷、重贷和再贷。这些用户的财务属性、信用属性和生活属性都表现出明显的负向倾向，所有成员都有逾期记录。其中，约有三分之一的用户偿债能力严重不足，部分用户也有参与娱乐输赢活动的习惯。在整体青年用户中，已婚有子女的比例相对较高，而逾期借贷型青年用户群体中，已婚无子女的比例较高。这表明，在婚后生活压力增加、家庭支出增多的情况下，用户更容易陷入财务困境，导致逾期借贷。同时，这个群体中个体工商户的比例较高，但他们的生产经营能力较差，说明部分青年用户在创业过程中可能面临经营困难，进而出现逾期借贷的情况。

这个群体的主要特征是信用意识不强，主要是还款能力不足的按揭购房人员，以及处于创业初期的小微商户。他们在财务上表现出较高的风险，借贷行为频繁且偿还能力较弱。生活中该群体可能由于家庭支出增加或收入不足而陷入困境，导致逾期情况发生。同时，该群体中有些人还会参与风险较高的娱乐活动，加剧了财务的不稳定性。总体来看，这一群体的逾期借贷行为是多种因素综合作用的结果，既有个人财务管理的问题，也有外部环境和生活压力的影响。

4.3 申请备用型用户

该群体占比略低于逾期借贷型用户，申请信贷的理由主要是为了备用金准备和业务被动营销。在财务属性方面，该群体的总资产分布广泛，大多数用户偿债能力较强且贷款占用情况表现正常。在性别比例上，该群体的分布为 1 : 2，与所有青年用户的性别比例分布（3 : 7）非常接近。在婚姻状况方面，已婚无子女和未婚人群占比较高，表明他们在家庭责任和经济压力相对较低的情况下，有较强的财务储备需求。职业属性上，该群体以办事人员和金融业从业者为主，这些用户通常具备 2~8 年的工作年限，拥有较好的经济基础。同时，该群体的大部分用户为本地人，居住状况以城区全款或自置为主，家庭人均月收入主要集中在 3000~5000 元之间。值得注意的是，所有用户均无逾期记录，但与手机银行业务的往来相对冷淡。

综合上述特征，该群体主要为小有资产的青年职工、经营良好的个体商户，以及金融机构内的营销推广人员。该群体的财务状况良好，具有较强的偿债能力和稳定的收入来源。尽管该群体的信贷需求主要是出于备用金的准备，但其整体财务健康度较高。在手机银行业务使用方面，可能由于他们目前的金融需求相对简单且习惯于传统的银行业务模式，这一群体存在较大的潜力待挖掘。

4.4 消费导向型用户

该群体与生产经营型用户的占比接近，申请信贷的理由偏向个人消费和消费升级。人口属性特征明显，可以描述为无性别、相对低龄、未婚类人群。未婚青年人群有更多的自由支配资金，倾向于进行消费性支出。在职业属性中，商业、服务业人员和办事人员等职业群体占比较高，说明该群体具备一定的消费能力，但收入水平相对较低。消费需求和消费欲望较为强烈，同时少部分青年用户在消费方面的风险意识不强，存在逾期记录。值得注意的是，该群体完全放弃自办经营实体，无人从事农、林、牧、渔业以及采矿业等传统行业。

相比于生产经营型群体，该群体的主要特征表现为本地年轻白领，开销较大，银户关系波动性强，是整个青年用户申贷行为的重要组成部分。他们的消费行为集中在日常生活和娱乐方面，表现出强烈的消费欲望和升级需求。同时，由于收入水平有限，该群体往往需要通过信贷来满足

消费需求，这也导致部分用户存在逾期记录。综上，该群体的申贷行为主要由消费需求驱动，具有较高的金融活跃度和较强的消费倾向。其信贷需求和还款能力呈现出较大的波动性，需要银行在产品设计和风险管理方面给予更多关注和支持。

4.5　初次申请型用户

该群体是青年用户中占比最低的群体，申请信贷的理由主要是为了尝试贷款的审慎式申请。在生活属性上，该群体的家庭人均月收入主要集中在 3000~8000 元之间，部分用户存在较强的房贷压力。职业和所涉及行业的分布较为宽泛，但工作年限都在 5 年以下，说明用户接触信贷的意愿有所偏差，有部分用户对于申贷行为的接受度仍有待提升。该群体中未婚占比高达 85.66%，其财务属性与信用属性互相对应，一方面财务压力相对较小，另一方面信用记录基本为白板，意味着该群体对于申贷行为的了解程度不高。

初次申请型用户主要为本地未婚青年，他们一般处于职业生涯起步阶段，信贷经验欠缺。他们的财务压力较小，但由于缺乏信贷经验，对申贷行为的接受度和了解程度不足。因此，这一群体需要增强对信贷行为的接受度和信贷知识的了解，以便更好地进行财务规划和管理。整体来看，这些用户的信贷需求较为谨慎，更倾向于尝试贷款并逐步了解信贷产品的功能和优势。在信贷市场中，这一群体具有较大的潜力，但需要通过有效的金融教育和支持来帮助他们更好地融入信贷体系，从而提升其信贷经验和财务管理能力。

5　结论与建议

本文基于某手机银行 APP 中的青年用户申贷信息大样本数据，将青年用户在 APP 中的申贷行为记录及其关联的历史数据进行匿名化和脱敏处理，构建了一个涵盖多维度的青年用户申贷画像标签体系。围绕青年用户这一申贷群体，通过深入挖掘申贷理由文本的主题内容，分析青年申贷用户的群体分类及申贷画像差异。研究发现，青年用户群体可划分为生产经营型、逾期借贷型、申请备用型、消费导向型和初次申请型这 5 种类别，不同青年用户群体申贷理由的主题和用户画像存在明显差异。基于此，建议以手机银行为核心的信贷业务尽量采取针对性的优化策略和管理手段，以增强手机银行用户的满意度，提高信贷业务经营能力。具体措施如下。

第一，针对生产经营型用户，银行应当提供定制化的产业链金融服务，以提高生产效率和降低经营风险。具体服务包括供应链金融、农业保险、智能化生产技术和数据分析工具的引入。同时针对特定产业，设计适合的金融产品，如农业机械贷款、农产品预售贷款、设备购置贷款和生产线升级贷款。整合供应链金融，提供上下游企业的协同贷款，减少资金流转中的断层，并通过智能化管理工具提升生产过程的效率，实现智慧农业和智能制造，优化生产管理水平。

第二，针对逾期借贷型用户，银行应当提供综合的财务规划和债务重组服务。通过设计债务再融资和债务重组方案，优化负债结构和还款计划，缓解用户还款压力。建立个人信用评分模型和逾期预警系统，提前发现逾期风险，并提供个性化的信用修复方案，通过逐步小额还款和债务修复贷款帮助用户重建良好的信用记录。此外，还可以提供灵活的还款计划和债务重组产品，确保用户能够在可承受的范围内逐步恢复财务健康。

第三，针对申请备用型用户，银行应当提供全面的理财规划和资产配置建议。结合资产配置模型和风险管理工具，制定个性化的投资组合和财务规划方案，以优化资产配置。推出高流动性、低风险的短期理财产品和备用资金贷款，满足青年用户在紧急情况下的资金需求。此外，根据青年用户的资产情况和财务目标，银行可以推荐定制化的理财方案，帮助用户在风险可控的情况下实现财富增值，提高资金使用效率和安全性。

第四，针对消费导向型用户，银行应当推出多样化的消费金融产品，包括教育贷款、旅游贷款、家装贷款等，以满足不同消费需求。分析青年用户的消费偏好和消费心理，设计个性化的消费理财方案，帮助青年用户合理规划支出。同时为其推荐合适的信用卡和分期付款产品，支持青年用户的消费决策。加强消费行为和风险意识的培养，引导青年用户进行理性消费和有效的风险管理，避免过度负债。

第五，银行应当为初次申请型用户提供全面的信贷知识普及和信用建设方案，包括信用报告解读、信用评分提升和信用保护等内容，帮助青年用户了解基本的信贷知识和信用管理技巧。简化贷款申请流程，提供在线信用教育模块，使青年用户能够方便快捷地获取必要的信贷信息。推出小额信贷产品，降低借贷门槛，帮助青年用户建立良好的信用记录和信用意识，提升其信用评级和借贷能力，为未来的金融活动奠定坚实的基础。

现阶段，手机银行所受理的贷款种类主要为信用贷款，对信用贷款与抵押贷款、质押贷款、担保贷款行为进行联合分析的数字平台仍不够完备。未来研究可以考虑将线上的手机银行 APP 申请数据和线下的非信用贷款申请数据共同纳入用户画像标签体系中去，形成一个更加全面、精准的青年用户画像标签体系，以支撑用户申贷行为画像建设的持续完善。

【参考文献】

［1］佘楷文，申宇，赵绍阳. 大数据对银行信贷行为的影响——来自数字社会信用平台的证据［J］. 经济研究，2024，59（3）：147-165.

［2］温博慧，刘雨菲，程朋媛. 数字金融对传统银行小微贷款影响的空间效应——基于非平衡空间计量模型的实证检验［J］. 国际金融研究，2022（3）：45-55.

［3］修晶. 促进农村青年创新创业的金融支持计划研究——以共青团组织小额贷款项目为例［J］. 中国青年社会科学，2016，35（4）：54-58.

［4］Snellinger A. ‘Every household will be a micro-enterprise’: a youth micro-loan scheme’s role in restructuring Nepal［J］. Children’s Geographies, 2018, 16(1): 66-79.

［5］Jang M, Nikou S, Kim S. Factors influencing users’ perceptions of digital platform indispensability: a comparative study of Korea and Finland［J］. Behavioral Sciences (Basel, Switzerland), 2024, 14(6): 502-502.

［6］Davis C S, Peter M,W A B, et al. Impact of service-based student loan repayment program on the primary care workforce［J］. Annals of Family Medicine, 2023, 21(4): 327-331.

［7］Jonghee L. Studies on insolvency prediction for young Korean debtor［J］. Journal of Korean Family Resource Management Association, 2019, 23(2): 99-115.

［8］Stefanizzi S, Lysova T. Usury vulnerability: measuring and modelling the phenomenon in Italy［J］. Quality & Quantity, 2024, 58(4): 3845-3862.

［9］Modestino S A, Sederberg R, Tuller L. Assessing the effectiveness of financial coaching: evidence from the boston youth credit building initiative［J］. Journal of Consumer Affairs, 2019, 53(4): 1825-1873.

[10] 陈添源，梅鑫. 多源数据融合的用户画像识别与推荐实证研究 [J]. 情报理论与实践，2024，47（4）：171–180.

[11] Hoang D T, Sheng Q Z, Emma W Z, et al.CupMar: a deep learning model for personalized news recommendation based on contextual user–profile and multi–aspect article representation [J].World Wide Web, 2022, 26(2): 713–732.

[12] Huo Z, Wang T, Fan Y, et al.Privacy–preserving global user profile construction through federated learning [J]. International Journal of Computational Science and Engineering, 2023, 26(2): 199–209.

[13] Yang D, Ji M, Lv Y, et al.Power user portrait model based on random forest [J]. IET Cyber–Physical Systems: Theory Applications, 2023, 9(1): 17–26.

[14] Deng W,Liang G, Yu C, et al. An early warning model of telecommunication network fraud based on user portrait [J]. Computers, Materials Continua, 2023, 75(1): 1561–1576.

[15] Anwar M S, Wang J, Ullah A, et al. User Profile Analysis for Enhancing QoE of 360 Panoramic Video in Virtual Reality Environment [C] //2018 International Conference on Virtual Reality and Visualization (ICVRV). IEEE, 2018: 106–111.

[16] 魏明珠，张海涛，刘雅姝，等. 多维属性融合的社交媒体高影响力人物画像研究 [J]. 图书情报知识，2019（5）：73–79，100.

[17] 王若佳，严承希，郭凤英，等. 基于用户画像的在线健康社区用户流失预测研究 [J]. 数据分析与知识发现，2022，6（Z1）：80–92.

[18] 杨阳，余维杰. 融合弹幕内容特征与行为特征的用户画像研究——以B站教学类视频为例 [J]. 情报科学，2022，40（12）：161–169.

[19] 李嘉兴，王晰巍，常颖，等. 基于移动终端日志的微信老年用户使用行为画像研究 [J]. 图书情报工作，2019，63（22）：31–40.

[20] 严炜炜，曹灿瑜. 多平台视角下用户知识交流主题挖掘与画像分析——以ChatGPT话题为例 [J]. 现代情报，2024，44（7）：47–59.

[21] Mesfer A, Abdullah A. Customer satisfaction analysis with Saudi Arabia mobile banking apps: a hybrid approach using text mining and predictive learning techniques [J]. Neural Computing and Applications, 2024, 36(11): 6005–6023.

[22] Saibaba S. Examining the determinants of mobile banking App continuance intention in India: an extension of the IS success model [J]. Journal of Internet Commerce, 2024, 23(1): 50–89.

[23] 卫梦洁，李静萍. 数字普惠金融对商业银行信用风险的影响——基于存款和贷款结构的联合视角 [J]. 经济问题，2024（2）：32–40.

[24] 王萍，龙秋易，蒋健. 基于BERTopic模型的移动办公类应用与用户的适应模式研究 [J]. 图书情报工作，2024，68（5）：110–120.

[25] 杨思洛，吴丽娟. 基于BERTopic模型的国外信息资源管理研究进展分析 [J]. 情报理论与实践，2024，47（2）：189–197.

[26] 刘洋，柳卓心，金昊，等. 基于BERTopic模型的用户层次化需求及动机分析——以抖音平台为例 [J]. 情报杂志，2023，42（12）：159–167.

[27] 迟国泰，潘明道，程砚秋. 基于综合判别能力的农户小额贷款信用评价模型 [J]. 管理评论，2015，27（6）：42–57.

[28] Hafez M. Examining the effect of consumption values on mobile banking adoption in Bangladesh: the moderating role of perceived security [J]. Kybernetes, 2023, 52(12): 6232–6250.

[29] 张凡，李渊，张龙耀. 农村普惠金融服务点对农户信贷行为的影响 [J]. 南开经济研究，2024，（2）：215–232.

[30] 李好奇，林华珍，黄福. 个人小额信贷风险影响因素研究 [J]. 数理统计与管理，2022，41（6）：

1105–1115.

［31］陈秋华，杨慧荣，崔恒建. 变量筛选后的个人信贷评分模型与统计学习［J］. 数理统计与管理，2020，39（2）：368–380.

［32］王卫军，宁致远，董昊，等. 基于语义相似关系的学科交叉主题识别方法［J］. 情报学报，2024，43（1）：34–47.

［33］Chao L. Optimization of negative sample selection for landslide susceptibility mapping based on machine learning using K–means–KNN algorithm［J］. Earth Science Informatics, 2023, 16(4): 4131–4152.

Research on Constructing Group Portrait of Youth Users' Loan Application Behavior in Mobile Banking APP

Dou Luyao

(National Science Library (Wuhan), Chinese Academy of Sciences, Wuhan 430071, China)

Abstract:［**Purpose/Significance**］This study aims to explore the thematic features of loan application reasons among young users, construct a user profiling label system, and systematically understand the patterns and preferences of young users regarding online mobile banking loan applications. This will assist in optimizing strategies for age-appropriate development of mobile banking platforms.［**Method/Process**］A total of 681,166 high-precision loan application records from the backend of a mobile banking app were collected. The BERTopic model was utilized to distill the main themes of loan application reasons. User profiling labels were extracted across five dimensions: financial attributes, demographic attributes, occupational attributes, lifestyle attributes, and credit attributes, based on the characteristics of loan application behavior. A user profiling identification label system was constructed in conjunction with bank user rating and scoring criteria. User profiling was achieved through K-means clustering to reveal group characteristics and conduct comparative analysis of differences.［**Results/Conclusion**］The study identified 30 themes covering six behavioral dimensions of loan reasons among young users: daily life, financial management, asset expenditure, personal interests, family life, and business operations. Based on these themes and attribute characteristics, young user loan applicants were segmented into five categories: business operation-oriented, overdue borrowing-oriented, standby application-oriented, consumption-oriented, and first-time application-oriented. Significant differences were found among these user groups, suggesting the need for targeted incentive measures to enhance user stickiness and satisfaction on mobile banking platforms.

Keywords: Mobile banking; Youth users; Loan application behavior; User portrait

（本文责编：任全娥）